# Offline Electron Seeding Validation - Update


**Caleb Fangmeier**
Ilya Kravchenko, Greg Snow

University of Nebraska - Lincoln


EGM Reco/Comm/HLT meeting — June 22, 2018

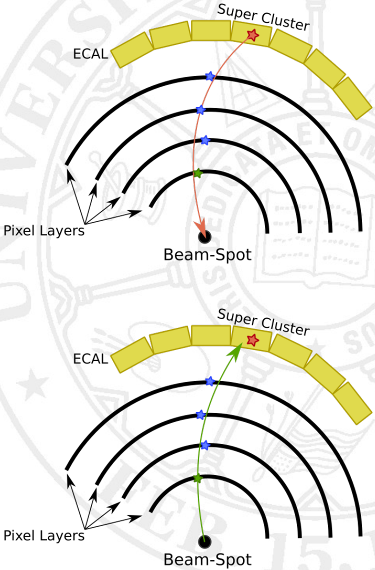
# Introduction

- Our goal is to study **seeding** for the **offline** GSF tracking with the **new pixel detector**.
- Specifically, we want to optimize the new pixel-matching scheme from HLT for use in off-line reconstruction.
- This Talk:
  - Define and demonstrate performance of a GSF-Track "Fake Rate" for:
    - Current offline (Legacy HLT) seeding method with default offline settings
    - New HLT seeding method with HLT settings[1]
    - New HLT seeding method with optimized-for-offline (aka `wide`) settings
  - Show efficiency for prompt prompt electrons specifically

[1] Note: In previous talks I've called this one `narrow`.

# N-Hit Electron Seeding

1. Using the beam spot, the SC position, and SC energy, propagate a path through the pixels.
2. Require the first hit to be within a $\delta\phi$ and $\delta z$ window. ($\delta\phi$ and $\delta R$ for FPIX)
3. $\delta z$ window for first hit is huge as SC and beam spot positions give very little information about $z$.
4. Forget the SC position, and propagate a new track based on the vertex and first hit positions, and the SC energy.
5. Progress one-by-one through the remaining hits in the seed and require each one fit within a specified window around the track.
6. Quit when all hits are matched, or a hit falls outside the window. No skipping is allowed.



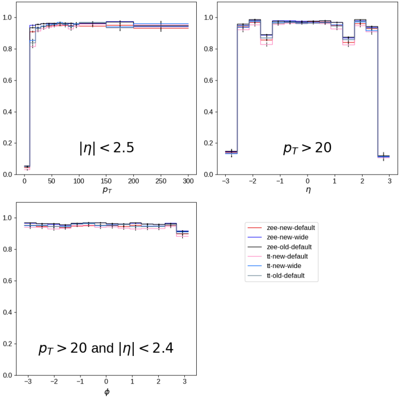
# Definitions

- **Sim-Track -** A track from a simulated electron both originating from the luminous region of CMS (beam-spot +- $5\sigma$) and having $|\eta| < 3.0$.
- **ECAL-Driven Seed -** A seed created via a matching procedure between Super-Clusters and General Tracking Seeds (Either from `ElectronSeedProducer` or `ElectronNHitSeedProducer`). Must have $HOE < 0.15$.
- **GSF Track -** A track from GSF-Tracking resulting from an **ECAL-Driven Seed**
- **GSF Tracking Efficiency -** The fraction of **Sim-Tracks** that have a matching **GSF Track** (based on $\Delta R$ matching)
- **GSF Tracking Purity -** The fraction of **GSF Tracks** that have a matching **Sim-Track**
- **GSF Tracking Fake Rate -** The fraction of nontruth-matched Super-Clusters which result in at least one **GSF Track**.

# Relative Performance - GSF Tracking Efficiency

- Figure shows GSF Tracking efficiency vs kinematic variables of the electron `SimTrack`
- Efficiency is more or less the same for both DY and $t\bar{t}$ environments and for both algorithms and working points.
- Largest (statistically significant) differences appear at low $p_T$ and in the barrel/endcap transition region.

GSF Tracking Efficiency

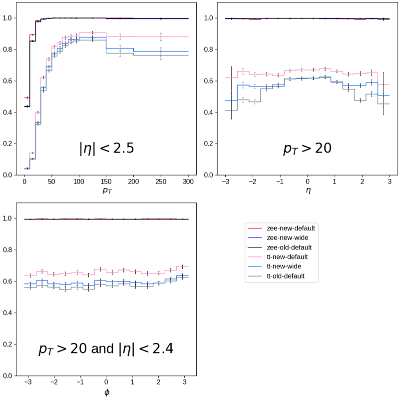



This and the following slide have been show before and are included for completeness

# Relative Performance - GSF Track Purity

- Figure shows GSF Tracking purity vs kinematic variables of the `GSFTrack`
- Clearly purity is affected by the higher fake environment in the $t\bar{t}$ sample.
- Note how the `narrow` working point of the new seeding (green) has significantly better purity than the `wide` working point or the old seeding.
- Purity loss at high $p_T$ is a feature of the shared-hits matching between `SimTracks` and `GSFTracks`.

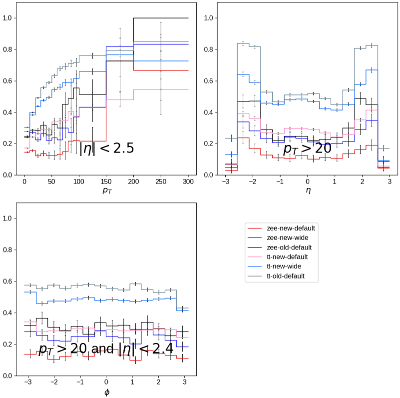
GSF Tracking Purity



# Relative Performance - GSF Tracking Fake Rate

- Figure shows GSF Tracking fake rate vs kinematic variables of the supercluster
- Supercluster must have $HOE < 0.15$, so fake are presumably from mostly photons or $\pi^0$
- There is a clear reduction in the fake rate with respect to the old method in both the `default` and `wide` working points.
- Seen in both $Z \to ee$ and $t\bar{t}$

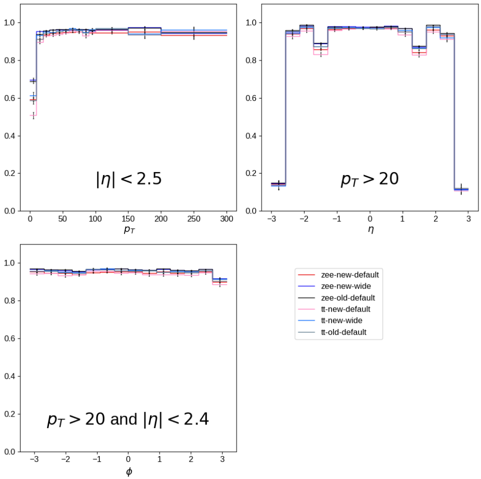
GSF Tracking Fake Rate



# Relative Performance - Prompt Efficiency

- The fraction of prompt electrons that match a GSF-Track
- Biggest improvements, again, happen at low $p_T$ and in the barrel/endcap transition region

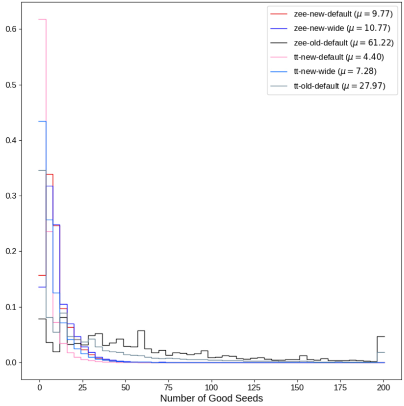
Prompt GSF Tracking Efficiency



# Relative Performance - Seed Multiplicity

- A single supercluster can potentially produce many seeds if it matches with many nearby tracks, however only one of these can be from the electron.
- Reducing the number of overall seeds while still producing *the* correct one is desirable from a computational perspective.
- The new seeding scheme (`wide` WP) reduces the number of seeds by a factor of 3.8 for $t\bar{t}$ and 5.6 for $Z \rightarrow ee$.

Number of Electron Seeds Per Event



# Overall Performance

Integrating over all tracks with $p_T > 20\text{GeV}$ and $\eta < 2.4$ yields the performance numbers below.

| Sample | Working Point | Efficiency | Purity | Fake Rate |
|---|---|---|---|---|
| $Z \rightarrow ee$ | new-default | $86.57 \pm 0.27\%$ | $90.58 \pm 0.28\%$ | $69.67 \pm 0.91\%$ |
| | new-wide | $87.90 \pm 0.27\%$ | $90.31 \pm 0.28\%$ | $76.45 \pm 0.91\%$ |
| | old-default | $87.95 \pm 0.27\%$ | $90.19 \pm 0.28\%$ | $79.40 \pm 0.91\%$ |
| $t\bar{t}$ | new-default | $86.95 \pm 0.77\%$ | $60.56 \pm 0.65\%$ | $31.40 \pm 0.53\%$ |
| | new-wide | $88.30 \pm 0.77\%$ | $54.35 \pm 0.61\%$ | $49.95 \pm 0.52\%$ |
| | old-default | $88.07 \pm 0.77\%$ | $52.25 \pm 0.60\%$ | $56.87 \pm 0.52\%$ |

- The HLT default settings (`narrow`) of the new pixel matching scheme yield non-trivially better purity at the loss of some efficiency with respect to both the old seeding and the `wide` working point.
- The `wide` working point of the new seeding matches the `old-seeding` within errors except for purity is $\approx 2\%$ better in the $t\bar{t}$ sample

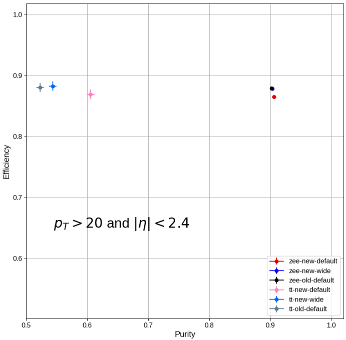
# Conclusions & Outlook

- The new seeding algorithm has been optimized to have better or comparable performance to the current Offline seeding method in all investigated metrics including
  - GSF Tracking Efficiency
  - GSF Tracking Purity
  - GSF Tracking Fake Rate
  - Number of Seeds
- Unless there are objections, propose to move forward with implementing the new algorithm as the default in the next available CMSSW release.

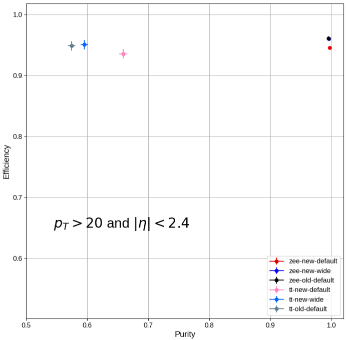
BACKUP

# Overall Performance

GSF Tracking Performance (Hit Matched)




GSF Tracking Performance ($\Delta R$ Matched)



# Matching Window Parameters

| | | narrow | default (HLT) | wide | extra-wide |
|---|---|---|---|---|---|
| Hit 1 | dPhiMaxHighEt | **0.025** | **0.05** | **0.1** | **0.15** |
| | dPhiMaxHighEtThres | 20.0 | 20.0 | 20.0 | 20.0 |
| | dPhiMaxLowEtGrad | -0.002 | -0.002 | -0.002 | -0.002 |
| | dRzMaxHighEt | 9999.0 | 9999.0 | 9999.0 | 9999.0 |
| | dRzMaxHighEtThres | 0.0 | 0.0 | 0.0 | 0.0 |
| | dRzMaxLowEtGrad | 0.0 | 0.0 | 0.0 | 0.0 |
| Hit 2 | dPhiMaxHighEt | **0.0015** | **0.003** | **0.006** | **0.009** |
| | dPhiMaxHighEtThres | 0.0 | 0.0 | 0.0 | 0.0 |
| | dPhiMaxLowEtGrad | 0.0 | 0.0 | 0.0 | 0.0 |
| | dRzMaxHighEt | **0.025** | **0.05** | **0.1** | **0.15** |
| | dRzMaxHighEtThres | 30.0 | 30.0 | 30.0 | 30.0 |
| | dRzMaxLowEtGrad | -0.002 | -0.002 | -0.002 | -0.002 |
| Hit 3+ | dPhiMaxHighEt | **0.0015** | **0.003** | **0.006** | **0.009** |
| | dPhiMaxHighEtThres | 0.0 | 0.0 | 0.0 | 0.0 |
| | dPhiMaxLowEtGrad | 0.0 | 0.0 | 0.0 | 0.0 |
| | dRzMaxHighEt | **0.025** | **0.05** | **0.1** | **0.15** |
| | dRzMaxHighEtThres | 30.0 | 30.0 | 30.0 | 30.0 |
| | dRzMaxLowEtGrad | -0.002 | -0.002 | -0.002 | -0.002 |

`NHit` Seeding window parameters. Bold designates modified values.

# Samples

- /ZToEE_NNPDF30_13TeV-powheg_M_120_200/RunIISummer17DRStdmix-NZSFlatPU28to62_92X_upgrade2017_realistic_v10-v1
- /TT_TuneCUETP8M2T4_13TeV-powheg-pythia8/RunIISummer17DRStdmix-NZSFlatPU28to62_92X_upgrade2017_realistic_v10-v2